1. Stanines
2. Norm referenced tests
3. Scale scores
4. Standardized and other formal assessments: Basic concepts

Stanines
An overview of stanines

Introduction
Using stanines as a measure for testing dates back to the 1940s. Stanine, literally means standard nine, giving us 9 bands of which the test scores are placed into. The normal distribution cure is a useful way to represent the stanines with each band having a width of 1/2 a standard deviation (excluding the 1st and 9th stanines at the outer edges).

Working out stanines for a set of test scores
It is possible to work out the stanine distributions for a set of test scores by doing the following:

1. Put all of the test scores into order from lowest to highest

2. Give the lowest 4% of test results a stanine 1, the next 7% will be allocated a stanine 2 and so forth... the following table provides the relationship between the percentage of test scores that are placed in each of the 9 stanine bands.

Stanine	1	2	3	4	5	6	7	8	9
Percentage in each stanine band	4%	7%	12%	17%	20%	17%	12%	7%	4%

Stanines

[missing_resource: http://assessments.lmi-usa.com/whytest/Images/BellCurve.jpg]

Source: http://assessments.lmi-usa.com/whytest/Staninesystem.asp

Interpreting Stanine Scores
Stanines can be useful, they provide a nice simple way of grouping students into fairly course but useful groups. They give a good representation of where a student is placed in relation to others in a group. The actual group the stanine score is representative of could be as small as the class of students sitting a test or as large as a representative sample of students that have been used to calculate expected norms of a test. The following scenarios are useful to work through in helping to build a clearer understanding of stanine scores. They are produced here with permission from Educational Measurement: Issues and Practice, Fall 1983 (John R. Hills - Florida State University).
Exercise:
A

Problem:

Mary is a sixth grader. She received a stanine score of zero on her standardized test in mathematics. This means that Mary's score was very low com pared to other sixth graders. Is that correct?

Solution:

No. Stanine scores are numbers from I to 9. There is no such thing as a stanine score of zero. A zero score reported as a stanine indicates that an error has been made.

Exercise:
B

Problem:

Bill received a stanine score of 5 on the same standardized mathematics test that Mary took. He is also in the sixth grade. The score of 5 means that Bill is doing average work in mathematics, and he would be at the 50th percentile for sixth graders. Is that correct?

Solution:

Yes and no. Stanine 5 is in the middle of the scale, and in that sense Bill got an average score on the test. However, each stanine represents a band of scores, not a specific score. The 5th stanine extends from the 40th to the 60th percentile. So Bill might be performing as low as the 40th percentile or as high as the 60th percentile but still receive a stanine of 5. However, because the stanine scale reflects a normal curve, the 40th percentile is usually only a few raw score points lower than the 60th percentile.

Exercise:
C

Problem:

Pedro received a stanine score of 6.5 on the mathematics test. This score should be inter preted as being midway between the sixth and seventh stanines.

Solution:

No. Stanines are represented by the single digit whole numbers, such as I, 2, and 3, never by numbers with decimal points. Except for the first and ninth sta nines, each stanine represents a narrow band of scores on the test. (The first and ninth stanines may be very wide in terms of raw score points. Each extends to the beginning or end of the test, however far that may be.) Thus, a stanine of 6.5 does not exist. Anyone who uses such a number for a stanine has made an error.

Exercise:
D

Problem:

Cindy is in the same class as Bill, Mary, and Pedro. On the mathematics test, she received a stanine score of9. Her mother wants to know just how high that score is-what percent of pupils perform less well than Cindy. Ms. Billingsley tells Cindy's mother that 96 percent of students in Cindy's grade performed less well than Cindy. Is this an accurate statement of Cindy's percentile rank?

Solution:

No. The ninth stanine is not the 96th percentile. The lower limit of the ninth sta nine is the 96th percentile, but the upper limit is plus infinity. Any performance above the 96th percentile is the 9th stanine. Cindy may have scored far above the 96th percentile and received a stanine of 9. The same is true at the other end of the scale for a stanine of 1.A person with a stanine score of 1 may be as high as the 4th percentile, or very much lower.

Exercise:
E

Problem:

Alfonso's stanine score is 7. Mr. Rivera is more familiar with standard scores than sta nines. He asked Ms. Billingsley how many standard deviations above the mean a stanine score of7 was. Ms. Billingsley imme diately responded, "One." Does Ms. Billingsley have a trick for remembering such things so well?

Solution:

Yes. Three easy landmarks for relat ing stanines to standard scores ·are the mean and plus and minus one standard deviation. The mean is in the middle of the fifth stanine. Plus one standard deviation is in the middle of the seventh stanine. Minus one standard deviation is in the middle of the third stanine.

Exercise:
F

Problem:

Mr. Rivera decided that Ms. Billingsley really knew her sta nines. So he pushed his luck and asked her what percent of students got stanine scores of7. Ms. Billingsley thought for a moment. Then she replied, "In a normal distribution, 12 per cent of the scores will be in the seventh stanine." Taken aback by the speed of her response, Mr. Rivera asked whether an other trick was involved. Was there?

Solution:

Yes. Ms. Billingsley used the Rule of Four. With stanines, a close approxima tion to the distribution of scores can be remembered as starting with 4 percent in either stanine 1 or 9, then adding 4 percent for the next stanine each time up to stanine 5 and then subtracting 4 percent for each to the end of the scale. Thus, the percent of the scores that are assigned I, 2, 3, ...9 are verycloseto4,8, 12, 16,20, 16, 12,8,and4, respectively. So Ms. Billingsley said to her self, "Four percent for stanine 9, 8 percent for stanine 8, and 12 percent for stanine 7." Then she had her

answer. She could have started with stanine 5, saying to herself, "Twenty percent in stanine 5, 16 percent in stanine 6, and 12 percent in stanine 7," reaching the same result.

Exercise:
G

Problem:

Mr. Tatnall overheard the conversation between Ms. Bill ingsley and Mr. Rivera and decided to contribute another guide. He suggested that sta nines were the same as deciles. So, he said, the first stanine would be the same as the first decile, the second stanine and the second decile would be equivalent, and so on. Is Mr. Tatnall correct?

Solution:

No. First, to be correct a decile is a point, not a range. The first decile is the score that separates the lowest scoring JO percent of scores from the highest scoring 90 percent, for example. The name for the lowest JO percent is the lowest tenth, or the first tenth, not the first decile. Beyond that, the first tenth is the lowest scoring JO per cent, but the first stanine is the lowest scor ing 4 percent, a much lower scoring group, on the average. In general, the only corres pondence between tenths of a distribution (or "deciles") and stanines is that tenths and stanines above 5 are high scoring and below 5 are low scoring. The differences between tenths and stanines reflect different as sumptions about the distribution of scores. Tenths are based on the assumption that scores have a rectangular or flat distribu tion. Stanines are based on the more realis tic assumption that scores are distributed normally.

Exercise:
H

Problem:

Mr. Rivera decided to ask one more question. He has found that most of his students receive the same stanine scores in the fifth grade that they got in the fourth grade or even the third grade. He concluded that they are not making much pro gress in school. ls that correct?

Solution:

No. Tests that use stanine scores refer these scores to students in a particular grade, not to students in general or to people in general. So a student who regularly receives stanine scores of 5 in a subject from year to year can be assumed to be making normal progress. He stays in the middle of the distribution. Another student who con tinually makes scores of stanine 7 stays about I standard deviation above the mean and makes normal progress also. Normal progress with stanines (or with percentiles or standard scores) is shown by earning the same score over time, not higher scores year by year.

Exercise:
I

Problem:

Mr. Tatnall asked what should he do about Patricia, who went down from the fifth stanine last year to the fourth stanine this year in reading comprehension? Should Mr. Tatnall be worried about this?

Solution:

No. Mr. Tatnall does not need to worry much about a change from one sta nine score to the adjacent stanine score. One question fewer correct could move a person one stanine down if his score was at the bottom of the range for that stanine. This is one of the problems with stanine scores. A person's performance can be anywhere in a range of scores but receive the same stanine. If Patricia scored at the lower edge of the fifth stanine, a trivial difference in performance could change her score to the next lower stanine.

Exercise:
J

Problem:

Mr. Rivera then asked about his student, Elena, whose stanine score in reading com prehension went up from the fourth stanine to the sixth sta nine. Is that big a difference important?

Solution:

Yes. When scores differ by two sta nines, we tend to think of there being a real difference, not an error of measurement. Other things being equal, for tests with satisfactory reliabilities (.90), such differen ces are expected to occur only about one time in ten. Therefore, differences that large deserve further investigation. Perhaps Elena has benefitted from some effective teaching, or she may have become more motivated, or she may have found more time to read, or something in her life that was impeding !}er progress may have been removed. A difference that large is unlikely to be an accident.

Norm referenced tests
Understanding what a norm referenced test is and what is important to know about the norms.

Introduction
Norm referenced tests enable the comparison of a test result with a wider sample of students, generally at the same year level. This wider sample is typically drawn from collecting data from a representative sample of the population. Two important ideas about norms are:

1. The norm reference data is only ever going to be as good as the sample it has been drawn from.
2. Norm reference data is specific to one point in time, and knowledge of what point in time the normed data represents is critical when analysing data.

1. The sample
The selection process for a sample is important in that once this data is collected, one hopes to be able to make reliable inferences about the larger population from which the sample has been drawn. In the development of tests for schools, it would be common to find a representative sample that draws upon the following for each year level:

1. A representative sample of schools in each of the decile rankings (1-10, or Low, Middle, High)
2. A representative number of students from the different groups of decile ranked schools.
3. A representative balance of gender
4. A representative sample of students from each of the main school size groups, for example: small schools, medium schools, large schools.

The size of the sample is also an important factor, and will range depending on the actual size of the population with which you are aiming to represent. The following site provides a high explanation of how you can go about calculating this kind of information,
http://en.wikipedia.org/wiki/Sample_size.

2. A point in time

Once the sample group is selected, the students will all be tested within the same time frame so that a set of norm reference data can be established. What is important here is to acknowledge that students continue to increase their knowledge and skills over time and so when one is wanting to compare a test result against the set of norm reference data, they are going to get the most accurate comparison when the point in time of the test is as closely matched to the point in time of the national norming study.

A common scenario for the development of school tests would be to produce one set of norms per year level. This norm reference data is often collected at the beginning of the year and for this reason, the normed data is a very good reference point when analysing test data from tests completed at the start of the year, a point in time that matches the norm reference data. When tests are undertaken at the end of the year, we have to be very careful to decide which reference data should be used that will best represent the expected level of academic achievement at the end of the year. Given a whole year of teaching and learning has occurred over the year, the start of year reference data is no longer the best set of data as it is based on student abilities from the start of the year. It is more accurate to be looking at the year above reference data as this point in time (start of the following year) is much closer to the students current level of ability, particularly given the Christmas holidays and less expected growth over these months without the regular teaching and learning programs occurring over a holiday period.

Scale scores
A description of scale scores

Introduction

A scaled score is the conversion of a student's raw score onto a common scale that allows for comparison between students and between different test scores from the same student. For this reason, the scale score is an excellent measure when looking at a students' progress over time. You are able to measure change from semester-to-semester, term-to-term, year-to-year of individual students or groups of students in a content area. A scale score provides a way to measure different tests that are targeted to students in different year levels and different levels of ability onto the same scale. There are two big picture ideas to understand about scale scores, these being:

1. A students scale score result can be placed onto the scale
2. The difficulty level of individual test items can be placed onto the same scale

1. Placing a student onto a scale

Because one scale is used for a test, for example a test in Mathematics. A students' raw score can be converted and placed onto the scale regardless of what year level the student is. So this means for example a year 3 student could complete a Maths test and have their score converted onto the Maths scale that has been created for this test. A year 10 student for example could complete a different Maths test, that has been calibrated onto the same Maths scale and then their result could also be converted from a raw score and placed onto the same scale. The likely scenario here, is that the year 3 student will be placed lower down on the scale where the mathematical demands on the student require less knowledge and skill and the year 10 student will be placed higher up on the scale where the demands of mathematical knowledge and skills are much higher. Unlike a stanine result that is based on norm referenced data from one point in time, the scale score result is just a scale score and can be placed onto a scale at any time of the year. For this reason, it is much easier to see where a student is at any time of the year and how they are progressing from a lower level of knowledge

and skills up to a great level of knowledge and skills as described by the scale that has been created for the test.

2. Placing items onto a scale

During the trial process of developing a test, the test developers are able to identify the difficulty of each item and place this on the same scale. The scale score of an item is a measure of the extent of knowledge and skills required from a student to be successful on the item. A difficult item has a high scale score because it requires more sophisticated skills and richer knowledge to be answered correctly than items lower on the scale. Different tests can be created that draw upon an easier set of questions, this test would then be well suited for students in the lower year levels. Likewise, a difficult test will draw upon a range of more difficult questions that are higher up on the scale. Having information from a test that provides not only a students score but also information about the type of knowledge and skills a student is able to complete brings a whole level of analysis that is not possible just from raw scores in themselves. This method of placing questions onto a scale comes from the Item Response Theory (IRT) http://en.wikipedia.org/wiki/Item_response_theory. A range of mathematical models can be used with this theory, the logistic and normal IRT models and the Rasch model are often used to calculate the placement of question and student scores onto the common equal interval scale for tests.

Standardized and other formal assessments: Basic concepts
Four scenarios of students experiencing standardized testing hurdles. Then, this module covers the basic concepts of standardized testing: when to use them, what context to use them in, how to look for the strengths and weaknesses of your students, and types of standardized tests.

Note: The primary author of this module is Dr. Rosemary Sutton.

Understanding standardized testing is very important for beginning teachers as K-12 teaching is increasingly influenced by the administration and results of standardized tests. Teachers also need to be able to help parents and students understand test results. Consider the following scenarios.

- *Vanessa, a newly licensed physical education teacher, is applying for a job at a middle school. During the job interview the principal asks how she would incorporate key sixth grade math skills into her PE and health classes as the sixth grade students in the previous year did not attain Adequate Yearly Progress in mathematics.*
- *Danielle, a first year science teacher in Ohio, is asked by Mr Volderwell, a recent immigrant from Turkey and the parent of a tenth grade son Marius, to help him understand test results. When Marius first arrived at school he took the Test of Cognitive Skills and scored on the eighty-fifth percentile whereas on the state Science Graduation test he took later in the school year he was classified as "proficient" .*
- *James, a third year elementary school teacher, attends a class in gifted education over summer as standardized tests from the previous year indicated that while overall his class did well in reading the top 20 per cent of his students did not learn as much as expected.*
- *Miguel, a 1st grade student, takes two tests in fall and the results indicate that his grade equivalent scores are 3.3 for reading and 3.0 for math. William's parents want him immediately promoted into the second grade arguing that the test results indicate that he already can read and do math at the 3rd grade level. Greg, a first grade teacher*

explains to William's parents that a grade equivalent score of 3.3 does not mean William can do third grade work.

Understanding standardized testing is difficult as there are numerous terms and concepts to master and recent changes in accountability under the *No Child Left Behind Act of 2001* (NCLB) have increased the complexity of the concepts and issues. In this chapter we focus on the information that beginning teachers need to know and start with some basic concepts.

Basic concepts

Standardized tests are created by a team—usually test experts from a commercial testing company who consult classroom teachers and university faculty—and are administered in standardized ways. Students not only respond to the same questions they also receive the same directions and have the same time limits. Explicit scoring criteria are used. Standardized tests are designed to be taken by many students within a state, province, or nation, and sometimes across nations. Teachers help administer some standardized tests and test manuals are provided that contain explicit details about the administration and scoring. For example, teachers may have to remove all the posters and charts from the classroom walls, read directions out loud to students using a script, and respond to student questions in a specific manner.

Criterion referenced standardized tests measure student performance against a specific standard or criterion. For example, newly hired firefighters in the Commonwealth of Massachusetts in the United States have to meet physical fitness standards by successfully completing a standardized physical fitness test that includes stair climbing, using a ladder, advancing a hose, and simulating a rescue through a doorway (Human Resources Division, nod.). Criterion referenced tests currently used in US schools are often tied to state content standards and provide information about what students can and cannot do. For example, one of the content standards for fourth grade reading in Kentucky is "Students will identify and describe the characteristics of fiction, nonfiction, poetry or plays" (Combined Curriculum Document Reading 4.1, 2006) and so a report on an individual student would indicate if the child can accomplish this skill. The report may

state that number or percentage of items that were successfully completed (e.g. 15 out of 20, i.e. 75 per cent) or include descriptions such as basic, proficient, or advanced which are based on decisions made about the percent of mastery necessary to be classified into these categories.

Norm referenced standardized tests report students' performance relative to others. For example, if a student scores on the seventy-second percentile in reading it means she outperforms 72 percent of the students who were included in the test's norm group. A norm group is a representative sample of students who completed the standardized test while it was being developed. For state tests the norm group is drawn from the state whereas for national tests the sample is drawn from the nation. Information about the norm groups is provided in a technical test manual that is not typically supplied to teachers but should be available from the person in charge of testing in the school district.

Reports from criterion and norm referenced tests provide different information. Imagine a nationalized mathematics test designed to basic test skills in second grade. If this test is norm referenced, and Alisha receives a report indicating that she scored in the eighty-fifth percentile this indicates that she scored better than 85 per cent of the students in the norm group who took the test previously. If this test is criterion-referenced Alisha's report may state that she mastered 65 per cent of the problems designed for her grade level. The relative percentage reported from the norm-referenced test provides information about Alisha's performance compared to other students whereas the criterion referenced test attempts to describe what Alisha or any student can or cannot do with respect to whatever the test is designed to measure. When planning instruction classroom teachers need to know what students can and cannot do so criterion referenced tests are typically more useful (Popham, 2004). The current standard-based accountability and NCLB rely predominantly on criterion based tests to assess attainment of content-based standards. Consequently the use of standardized norm referenced tests in schools has diminished and is largely limited to diagnosis and placement of children with specific cognitive disabilities or exceptional abilities (Haertel & Herman, 2005).

Some recent standardized tests can incorporate both criterion-referenced and norm referenced elements in to the same test (Linn & Miller, 2005). That is, the test results not only provide information on mastery of a content standard but also the percentage of students who attained that level of mastery.

Standardized tests can be high stakes i.e. performance on the test has important consequences. These consequences can be for students, e.g. passing a high school graduation test is required in order to obtain a diploma or passing PRAXIS II is a prerequisite to gain a teacher license. These consequences can be for schools, e.g. under NCLB an increasing percentage of students in every school must reach proficiency in math and reading each year. Consequences for schools who fail to achieve these gains include reduced funding and restructuring of the school building. Under NCLB, the consequences are designed to be for the schools not individual students (Popham, 2005) and their test results may not accurately reflect what they know because students may not try hard when the tests have low stakes for them (Wise & DeMars, 2005).

Uses of standardized tests

Standardized tests are used for a variety of reasons and the same test is sometimes used for multiple purposes.

Assessing students' progress in a wider context

Well-designed teacher assessments provide crucial information about each student's achievement in the classroom. However, teachers vary in the types of assessment they use so teacher assessments do not usually provide information on how students' achievement compares to externally established criteria. Consider two eighth grade students, Brian and Joshua, who received As in their middle school math classes. However, on the standardized norm referenced math test Brian scored in the fiftieth percentile whereas Joshua scored in the ninetieth percentile. This information is important to Brian and Joshua, their parents, and the school

personnel. Likewise, two third grade students could both receive Cs on their report card in reading but one may pass 25 per cent and the other 65 per cent of the items on the Criterion Referenced State Test.

There are many reasons that students' performance on teacher assessments and standardized assessments may differ. Students may perform lower on the standardized assessment because their teachers have easy grading criteria, or there is poor alignment between the content they were taught and that on the standardized test, or they are unfamiliar with the type of items on the standardized tests, or they have test anxiety, or they were sick on the day of the test. Students may perform higher on the standardized test than on classroom assessments because their teachers have hard grading criteria, or the student does not work consistently in class (e.g. does not turn in homework) but will focus on a standardized test, or the student is adept at the multiple choice items on the standardized tests but not at the variety of constructed response and performance items the teacher uses. We should always be very cautious about drawing inferences from one kind of assessment.

In some states, standardized achievement tests are required for home-schooled students in order to provide parents and state officials information about the students' achievement in a wider context. For example, in New York home-schooled students must take an approved standardized test every other year in grades four through eight and every year in grades nine through twelve. These tests must be administered in a standardized manner and the results filed with the Superintendent of the local school district. If a student does not take the tests or scores below the thirty-third percentile the home schooling program may be placed on probation (New York State Education Department, 2005).

Diagnosing student's strengths and weaknesses

Standardized tests, along with interviews, classroom observations, medical examinations, and school records are used to help diagnose students' strengths and weaknesses. Often the standardized tests used for this purpose are administered individually to determine if the child has a disability. For

example, if a kindergarten child is having trouble with oral communication, a standardized language development test could be administered to determine if there are difficulties with understanding the meaning of words or sentence structures, noticing sound differences in similar words, or articulating words correctly (Peirangelo & Guiliani, 2002). It would also be important to determine if the child was a recent immigrant, had a hearing impairment or mental retardation. The diagnosis of learning disabilities typically involves the administration of at least two types of standardized tests—an aptitude test to assess general cognitive functioning and an achievement test to assess knowledge of specific content areas (Peirangelo & Guiliani, 2006). We discuss the difference between aptitude and achievement tests later in this chapter.

Selecting students for specific programs

Standardized tests are often used to select students for specific programs. For example, the SAT (Scholastic Assessment Test) and ACT (American College Test) are norm referenced tests used to help determine if high school students are admitted to selective colleges. Norm referenced standardized tests are also used, among other criteria, to determine if students are eligible for special education or gifted and talented programs. Criterion referenced tests are used to determine which students are eligible for promotion to the next grade or graduation from high school. Schools that place students in ability groups including high school college preparation, academic, or vocational programs may also use norm referenced or criterion referenced standardized tests. When standardized tests are used as an essential criteria for placement they are obviously high stakes for students.

Assisting teachers' planning

Norm referenced and criterion referenced standardized tests, among other sources of information about students, can help teachers make decisions about their instruction. For example, if a social studies teacher learns that most of the students did very well on a norm referenced reading test

administered early in the school year he may adapt his instruction and use additional primary sources. A reading teacher after reviewing the poor end-of-the-year criterion referenced standardized reading test results may decide that next year she will modify the techniques she uses. A biology teacher may decide that she needs to spend more time on genetics as her students scored poorly on that section of the standardized criterion referenced science test. These are examples of assessment for learning which involves data-based decision making. It can be difficult for beginning teachers to learn to use standardized test information appropriately, understanding that test scores are important information but also remembering that there are multiple reasons for students' performance on a test.

Accountability

Standardized tests results are increasingly used to hold teachers and administrators accountable for students' learning. Prior to 2002, many States required public dissemination of students' progress but under NCLB school districts in all states are required to send report cards to parents and the public that include results of standardized tests for each school. Providing information about students' standardized tests is not new as newspapers began printing summaries of students' test results within school districts in the 1970s and 1980s (Popham, 2005). However, public accountability of schools and teachers has been increasing in the US and many other countries and this increased accountability impacts the public perception and work of all teachers including those teaching in subjects or grade levels not being tested.

For example, Erin, a middle school social studies teacher, said:

"“As a teacher in a 'non-testing' subject area, I spend substantial instructional time suporting the standardized testing requirements. For example, our school has instituted 'word of the day', which encourages teachers to use, define, and incorporate terminology often used in the tests (e.g. "compare", "oxymoron" etc.). I use the terms in my class as often as possible and incorporate them into written assignments. I also often use test questions of similar formats to the standardized tests in my own subject

assessments (e.g. multiple choice questions with double negatives, short answer and extended response questions) as I believe that practice in the test question formats will help students be more successful in those subjects that are being assessed." "

Accountability and standardized testing are two components of Standards Based Reform in Education that was initiated in the USA in 1980s. The two other components are academic content standards which are described later in this chapter and teacher quality which was discussed in [link]Chapter 1.

Types of standardized tests

Achievement tests

Summarizing the past: K-12 achievement tests are designed to assess what students have learned in a specific content area. These tests include those specifically designed by states to access mastery of state academic content standards (see more details below) as well as general tests such as the California Achievement Tests, The Comprehensive Tests of Basic Skills, Iowa Tests of Basic Skills, Metropolitan Achievement Tests, and the Stanford Achievement Tests. These general tests are designed to be used across the nation and so will not be as closely aligned with state content standards as specifically designed tests. Some states and Canadian Provinces use specifically designed tests to assess attainment of content standards and also a general achievement test to provide normative information.

Standardized achievement tests are designed to be used for students in kindergarten though high school. For young children questions are presented orally, and students may respond by pointing to pictures, and the subtests are often not timed. For example, on the Iowa Test of Basic Skills (http://www.riverpub.com/) designed for students are young as kindergarten the vocabulary test assesses listening vocabulary. The teacher reads a word and may also read a sentence containing the word. Students are then asked to choose one of three pictorial response options.

Achievement tests are used as one criterion for obtaining a license in a variety of professions including nursing, physical therapy, and social work, accounting, and law. Their use in teacher education is recent and is part of the increased accountability of public education and most States require that teacher education students take achievement tests in order to obtain a teaching license. For those seeking middle school and high school licensure these are tests are in the content area of the major or minor (e.g. mathematics, social studies); for those seeking licenses in early childhood and elementary the tests focus on knowledge needed to teach students of specific grade levels. The most commonly used tests, the PRAXIS series, tests I and II, developed by Educational Testing Service, include three types of tests (www.ets.org):

- Subject Assessments, these test on general and subject-specific teaching skills and knowledge. They include both multiple-choice and constructed-response test items.
- Principles of Learning and Teaching (PLT) Tests assess general pedagogical knowledge at four grade levels: Early Childhood, K-6, 5-9, and 7-12. These tests are based on case studies and include constructed-response and multiple-choice items. Much of the content in this textbook is relevant to the PLT tests.
- Teaching Foundations Tests assess pedagogy in five areas: multi-subject (elementary), English, Language Arts, Mathematics, Science, and Social Science.

These tests include constructed-response and multiple-choice items which tests teacher education students. The scores needed in order to pass each test vary and are determined by each state.

Diagnostic tests

Profiling skills and abilities: Some standardized tests are designed to diagnose strengths and weaknesses in skills, typically reading or mathematics skills. For example, an elementary school child may have difficult in reading and one or more diagnostic tests would provide detailed information about three components: (1) word recognition, which includes

phonological awareness (pronunciation), decoding, and spelling; (2) comprehension which includes vocabulary as well as reading and listening comprehension, and (3) fluency (Joshi 2003). Diagnostic tests are often administered individually by school psychologists, following standardized procedures. The examiner typically records not only the results on each question but also observations of the child's behavior such as distractibility or frustration. The results from the diagnostic standardized tests are used in conjunction with classroom observations, school and medical records, as well as interviews with teachers, parents and students to produce a profile of the student's skills and abilities, and where appropriate diagnose a learning disability.

Aptitude tests

Predicting the future: Aptitude tests, like achievement tests, measure what students have learned, but rather than focusing on specific subject matter learned in school (e.g. math, science, English or social studies), the test items focus on verbal, quantitative, problem solving abilities that are learned in school or in the general culture (Linn & Miller, 2005). These tests are typically shorter than achievement tests and can be useful in predicting general school achievement. If the purpose of using a test is to predict success in a specific subject (e.g. language arts) the best prediction is past achievement in language arts and so scores on a language arts achievement test would be useful. However when the predictions are more general (e.g. success in college) aptitude tests are often used. According to the test developers, both the ACT and SAT Reasoning tests, used to predict success in college, assess general educational development and reasoning, analysis and problem solving as well as questions on mathematics, reading and writing (http://www.collegeboard.com; http://www.act.org/). The SAT Subject Tests that focus on mastery of specific subjects like English, history, mathematics, science, and language are used by some colleges as entrance criteria and are more appropriately classified as achievement tests than aptitude tests even though they are used to predict the future.

Tests designed to assess general learning ability have traditionally been called Intelligence Tests but are now often called learning ability tests,

cognitive ability tests, scholastic aptitude tests, or school ability tests. The shift in terminology reflects the extensive controversy over the meaning of the term intelligence and that its traditional use was associated with inherited capacity (Linn & Miller 2005). The more current terms emphasize that tests measure developed ability in learning not innate capacity. The Cognitive Abilities Test assesses K-12 students' abilities to reason with words, quantitative concepts, and nonverbal (spatial) pictures. The Woodcock Johnson III contains cognitive abilities tests as well as achievement tests for ages 2 to 90 years (http://www.riverpub.com).